Article

Journal of Advanced Academics
2015, Vol. 26(4) 274–293

DOI: 10.1177/1932202X15603365
joaa.sagepub.com

Effects of Participation in a Post-Secondary Honors Program With Covariate Adjustment Using Propensity Score

Scott R. Furtwengler[1,2]

Abstract

The present study sought to determine the extent to which participation in a post-secondary honors program affected academic achievement. Archival data were collected on three cohorts of high-achieving students at a large public university. Propensity scores were calculated on factors predicting participation in honors and used as the covariate. The Johnson–Neyman technique was used to determine the values of the covariate on which the groups differed. The effect of participating in honors was greater for smaller propensity scores. The findings offer a positive outcome associated with participation for students at the lower end of the propensity score continuum, providing evidence to conclude that such programs are beneficial to a subset of high-achieving students.

Keywords

high achievers, honors programs, propensity score, Johnson–Neyman technique

Introduction

Colleges and universities are being held increasingly accountable for student success as measured by persistence, retention, time to graduation, and completion. Referred to as *performance funding* (Burke & Modaressi, 2001), some states are considering or

[1]University of Houston, TX, USA
[2]Brazosport College, Lake Jackson, TX, USA

Corresponding Author:
Scott R. Furtwengler, Office of Institutional Research, Brazosport College, 500 College Drive, Lake Jackson, TX 77566, USA.
Email: scott.furtwengler@brazosport.edu

are now using these metrics to determine and to justify the allocation of funds to their public schools. Because of this, questions have arisen as to whether supporting post-secondary honors programs is the most appropriate use of funds and resources for the special population of high-achieving learners who attend institutions of higher education (K. Campbell, 2005). Administrators are being challenged as to whether or not honors programs are effective in meeting the learning needs of high-achieving learners, with some administrators and faculty favoring elimination of institutional support of honors programs amid claims that such programs are elitist (McDermott, 1989; Pehlke, 2003; Selingo, 2002). Others argue, however, that honors programs provide the optimal environment for high-achieving learners and that the best practices gleaned from such progressive programs "trickle down" to traditional curricula (Giazzoni & Hilberg, 2009; Mihelich, Storrs, & Pellett, 2007; Moritz, 2011; Rinn & Cobane, 2009; Shushok, 2002; Sievert, 1972; Spurrier, 2009; Weiner, 2009).

More quantitative research of high-achieving students within the context of post-secondary honors programs is necessary to increase validity and generalizability of the results and conclusions regarding the effects of participating in honors and on the factors that influence high-achieving students to participate in honors. Building on previous, yet limited, research, the current study served to complement the current body of research by conducting an observational study analyzing recent, archival data on high-achieving students at a large, public research institution using a quasi-experimental design.

Achterberg (2005) argued that high-achieving learners need to have a classroom or learning environment that is differentiated from the conventional curriculum because such students "respond differently and learn better with other highly motivated, 'like-minded' students" (p. 80). Seifert, Pacarella, Colangelo, and Assouline (2007) concluded that participation in honors programs significantly affects students' undergraduate experience regarding academic achievement and cognitive development, but hastened to point out that the effects may not be as global as some proponents of honors education have suggested. Despite the data supporting such differences between honors and typical curricula, the researchers offered no factors that would account for such differences and, based on the nature of the study design, no correlation between participation in honors and the measured outcomes could be inferred. Seifert et al. noted that such research would be valuable so that the strategies that are found to be effective in an honors environment might be applied to traditional learning environments.

Research findings support the phenomenon that high-achieving students opt out of post-secondary honors programs because they perceive participation in such programs will result in lower academic achievement as measured by grade point average (GPA; Chase, Hemmeter, & Griffin, 1971; Hartleroad, 2005). Moon (2012) discovered that eligible students who forgo the honors experience tend to be male, tend to identify as an underrepresented race or ethnicity, and are first-generation college students. These students reported that they perceived honors as extra work without adequate benefit, were concerned about overextending their time commitments and increased stress levels, lacked a clear understanding about the program, and lacked academic self-efficacy.

Furthermore, other researchers found that a majority of high-achieving, low-income students did not apply to selective post-secondary programs because they and their parents were unaware of the scholarships that were available and that the students lacked a "sense of belonging" to the program (Hoxby & Avery, 2012). Such perceptions can result in eligible high-achieving learners forgoing an opportunity to engage in learning environments potentially better suited to their particular academic needs and goals (Achterberg, 2005).

Literature Review

A review of the literature revealed a modest body of research exploring between-group differences among high-achieving students who participate in post-secondary honors programs and those who do not. Several researchers have reported that honors programs provide an optimal environment for high-achieving learners but did not provide evidence of a causal relationship between participation in honors and academic outcomes (Giazzoni & Hilberg, 2009; Mihelich et al., 2007; Moritz, 2011; Rinn & Cobane, 2009; Shushok, 2002; Sievert, 1972; Spurrier, 2009; Weiner, 2009).

Although several studies have attempted to explore quantitative comparisons in academic performance between honors and non-honors students, they have been limited by small sample sizes, controlling for confounding variables, and problems in generalizability. Shushok (2006) noted the lack of literature using scientific methodologies to explore the effects of participating in an honors program. A small body of research focuses on whether a statistically significant difference exists in academic performance between students who participate in honors programs and those with similar academic ability who choose to opt out of such experiences. Of these, a study by Teske and Etheridge (2010), limited to first-year students, focused on information and communication literacy and found that 44.2% of non-honors freshmen met or exceeded the national average score of 149.8 on a test of information literacy. The honors freshmen in their study passed at double the rate (88.9%) with an average score of 179.2. It is difficult, however, to attribute these differences in performance on this instrument to participation in honors or to other factors.

Ogilvie and Reza (2009) reported that in 8 of 10 measures, students in an honors section of a business course scored significantly higher than non-honors students in an identical non-honors section. This study was limited, however, to the participants' experience and performance in only one section of an honors business course and an identical section offered to non-honors students. The relatively small sample consisted of 43 non-honors students and 19 honors students, making it challenging to attribute any effect to participation in honors.

Addressing concerns that participation in honors programs may result in a "penalty" affecting GPA, Chase et al. (1971) reported that participating in an honors curriculum did not negatively affect the academic achievement outcomes of high-achieving students and that there was a statistically significant difference in mean GPA between the honors students' non-honors coursework and honors coursework, with students earning a higher mean level GPA in the honors coursework. Chase et al.'s study was

limited to honors students, comparing their performance in honors courses with their performance in traditional (non-honors) courses over 4 years. Hartleroad's (2005) study compared the academic performance of honors and non-honors students, but it was limited in scope to first-year female engineering students. The researchers' findings indicated a statistically significant difference in mean GPA with honors students outperforming non-honors students. Similar to other studies, it was clear that honors students demonstrated better performance in honors courses than non-honors courses and better performance than their non-honors peers, but the conclusion that these differences were an effect of participation in honors was not supported.

Further empirical research provides evidence for positive and significant relationships between participating in honors and academic achievement, even when controlling for confounding factors and covariates, although Astin (1978, 1993) reported mixed results. Cosgrove (2004) reported that among students who graduate within 5 years at a 4-year institution, honors students who completed the honors program earned a mean cumulative GPA of 3.71, honors students who partially completed the program earned a GPA of 3.48, and non-honors students completing the traditional curriculum earned a mean cumulative GPA of 3.36. These differences in mean cumulative GPA, however, were not examined for significance. Rinn (2007) reported that honors students demonstrate higher academic achievement and higher academic self-concept than non-honors students, even after controlling for Scholastic Aptitude Test (SAT) scores. Whether or not higher scores on measures of self-concept were affected by participating in honors or whether the decision to participate in honors was mediated by self-concept was not communicated.

Recent quantitative research on post-secondary honors programs has utilized propensity score methods. Three studies comparing outcomes between honors and non-honors students used SAT scores and either high school GPA or class rank as observable characteristics (Keller & Lacy, 2013; Shushok, 2006; Slavin, Coladarci, & Pratt, 2008). The researchers constructed subsamples of non-honors students whose past academic achievements and academic potential approximated those same characteristics among honors students. Using SAT scores and high school class rank comparable with honors students, Slavin et al. (2008) identified a group of high-achieving non-honors students that approximated the academic profile of honors students at the University of Maine. Statistically controlling for these observed variables, the researchers compared the two groups to determine whether a relationship existed between honors participation and retention and graduation rates. Their findings indicated that 1-year retention rates were significantly higher for honors students than for non-honors students with 94% of the 2006 honors students returning as compared with 85% of non-honors students. In addition, using logistic regression the researchers reported that an honors student was 3.1 times more likely to return after 1 year than a non-honors student. Their findings, however, did not indicate a significant difference in the 4-year graduation rates for the entering 2002 cohort, with 64% of honors students and 60% of non-honors students graduating in 4 years.

Using a propensity score caliper matching method to define the acceptable level of comparability, Shushok (2006) selected a non-honors group the same size ($n = 86$) and

with comparable characteristics as an honors group (n = 86) at the University of Maryland, College Park. Using this matching technique, the researcher deemed the two groups as comparable if their mean scores on observable measures (e.g., GPA and SAT) were within 0.15 of a standard deviation of one another. The matching process resulted in honors and non-honors groups whose matched pairs had "identical" mean high school GPAs and mean SAT scores. The two groups had equivalent percentages of females, percentages living on campus, and percentages of ethnicities. Findings indicated that honors students had significantly higher cumulative GPAs at the end of the first year (honors, M = 3.41 and non-honors, M = 3.18) and higher 1-year retention rates (honors 97% and non-honors 90%). Similar to the Slavin et al. (2008) study, however, these differences tended to disappear by the fourth year.

To better estimate the unique contribution of an honors program to retention and graduation outcomes, Keller and Lacy (2013) applied propensity score analysis to control for the effects of honors students' previous academic achievement and other background characteristics versus the effects of program participation. The study statistically adjusted for a wider range of confounding characteristics than had been typical in previous research. In addition to academic achievement in high school, they included data on ethnic status, gender, in-state/out-of-state origin, first-generation college attendance, and academic unit at entry. Academic achievement was measured using the State of Colorado's college admission index, which is a composite encompassing high school grades or class rank and standardized test scores (American College Testing [ACT] or SAT). Adjusting for the influence of these background factors, the researchers' results indicated that the effects of the program itself were more modest than indicated by the raw comparison, but those effects were still found to be relatively large and statistically significant. They found the rate of 5-year graduation for honors students was 81.9% versus the 69.6% that would have been expected among these students had they not participated in the program. The researchers noted that this is a much smaller gain for honors participation than would be implied by a comparison with the unadjusted honors graduation rate of 57.6% and demonstrated the potential importance of an adjusted comparison in describing retention and graduation outcomes among honors participants. The adjusted comparisons of effects on 5- and 6-year graduation, however, show much larger differences than on 4-year graduation. These results seem to contradict the notion of a penalty associated with honors coursework.

Much of the between-group empirical research exploring differences between honors and non-honors students has been qualitative (Blythe, 2004; Butler, Pryor, & Marti, 2004; Castro-Johnson & Wang, 2003; Hébert & McBee, 2007; Morris, 2008), several researchers have conducted quantitative studies, as noted above, yet much of this research on the effects of participation in honors has been limited to within-group studies (Cosgrove, 2004; Rice, Leever, Christopher, & Porter, 2006; Rinn, 2005; Siegle, Rubenstein, Pollard, & Romey, 2010; Wurst, Smarkola, & Gaffney, 2008). There are a substantial number of high-achieving students, however, who choose to forgo such an undergraduate experience or are simply unaware that such learning environments are an option for them. The body of literature exploring differences between

honors and non-honors students is fairly clear in drawing distinctions between the two groups on several measures. Research determining whether those differences may be attributable to participation in an honors program, and to what degree, between honors and non-honors students, however, is lacking. More research is necessary to determine the effect of an honors education and, if a positive or negative effect associated with participation, that high-achieving college-going students are aware of the associations and their options. The present study sought to determine the magnitude of effect on participation in honors as measured by cumulative GPA. Cumulative GPA has been found to be related to personality and motivation (Jaramillo & Spector, 2004), achievement striving (Bacon & Bean, 2006), individual learning (Karakaya, Ainscough, & Chopoorian, 2001), academic performance (Borde, 1998), team learning (Hite, McIntyre, & Lynch, 2001), and cheating behavior (Chapman, Davis, Toy, & Wright, 2004). Furthermore, GPA has the potential to account for nearly half the variance in educational research models (Bloom, 1976). Bacon and Bean (2006) found 4-year cumulative GPA reliability to be high (r = .94), a finding substantially higher than previous research by Roth and Clark (1998), in which a reliability coefficient of r = .84 was found, and that GPA reliability increased over a 4-year period. The researchers also found that cumulative GPA had greater predictive validity than content-limited (i.e., specific academic programs) GPAs, unless the research context was limited to seniors, in which context content-limited GPAs were more valid (Bacon & Bean, 2006). Based on previous research, the researcher hypothesized that students participating in a post-secondary honors program would score higher on measures of cumulative GPA than their non-honors counterparts and that part of the variance in performance could be explained by participation in honors.

Purpose

The present study was driven by the following research question:

> **Research Question 1:** Does participation in a post-secondary honors program bear a significant relationship to academic achievement as measured by cumulative GPA by high-achieving students?

The study analyzed archival data for three cohorts of entering freshman at a large, 4-year university to determine if indeed a correlation exists between participation in the university's honors college and academic achievement as measured by cumulative GPA. The purpose of the present study is to determine the estimated level of effect for participation in a post-secondary honors program on academic achievement.

To test whether a discrepancy exists between high-achieving students' academic performance, differences in mean level cumulative GPA were measured between students who participated in a post-secondary honors program and those who did not at a large, 4-year research university. Archival data were collected and analyzed to determine whether those who participated in the post-secondary honors program earned a higher cumulative GPA than those in the traditional curriculum. The researcher hypothesized

that there is a significant relationship between participation in honors and academic achievement, with learners in honors outperforming their non-honors peers on the measure of cumulative GPA.

By utilizing propensity score analysis, the study determined whether a significant relationship exists between participation in a post-secondary honors program and academic achievement. This will be useful to students and administrators in drawing conclusions regarding the effect of, respectively, participating in and supporting such post-secondary programs. There is a gap in the literature with regard to demonstrating strong correlations between participation in honors and academic outcomes. The present study will contribute to the gap of research in these areas.

Method

The study, an observational study, used archival data collected from three entire cohorts: 2009, 2010, and 2011. For the three cohorts, first-year cumulative GPA was collected at the end of the spring semester. Data were screened and cleaned to remove students with no reported high school percentile ranking (HSPR), SAT, or GPA. All part-time cases were removed. Analysis for outliers on measures of the observable variables SAT total score and HSPR were removed.

There are four types of propensity score analysis methods: matching on the propensity score, stratification on the propensity score, inverse probability of treatment weighting using the propensity score, and covariate adjustment using the propensity score. The current study uses the covariate adjustment method (Austin, 2011). Propensity scores were calculated as a method for better estimation of treatment effect on the criterion variable of cumulative GPA between the groups (non-honors = 0, honors = 1). The propensity score is defined as the probability of receiving treatment based on measured covariates (Thoemmes & Kim, 2011): $e(x) = P(\mathrm{Z} = 1 \mid X)$, where $e(x)$ is the propensity score, P is the probability, $Z = 1$ indicates receipt of treatment with values 0 for control or comparison group (i.e., non-honors) and 1 for treatment group (i.e., honors), and X is the observable characteristics or covariates (in this study, SAT scores, HSPR, gender, and ethnicity). That is to say, the propensity score refers to the likelihood that a participant will matriculate into the treatment group based on the observable covariates, for example, previous performance on a continuous measure or measures and demographic characteristics.

A propensity score for each student to determine predicted probability for participation in the honors group was calculated using four observable characteristics, including SAT total score, HSPR, gender, and ethnicity on the dichotomous variable of group (non-honors = 0, honors = 1). The resulting propensity score was then used as the covariate. Because the groups may have differed on a variety of characteristics, a quasi-experimental design demonstrated that the treatment (honors) and comparison (non-honors) groups were equivalent on observable characteristics (What Works Clearinghouse, 2011) to support external validity.

Factors comprising the pre-treatment covariate (propensity score) were selected based on theoretical frameworks, previous research findings, and conventional criteria

for admission to honors. This was a critical step as the integrity of the analysis was dependent on its utility in inferring a relationship between participating in honors and students' GPA based on the researcher's selection of appropriate characteristics to include in the calculation of the propensity score (Thoemmes & Kim, 2011). Demographic characteristics, such as gender, ethnicity, or age, are appropriate, yet not sufficient (Shadish, Clark, & Steiner, 2008). The inclusion of a measure of high school academic performance and standardized test scores provide additional metrics on which to match students on similar variables (Keller & Lacy, 2013; Shushok, 2006; Slavin et al., 2008; Tam & Sukatme, 2003, 2004). The researcher used four observable covariates on which propensity matching was based: gender, ethnicity, HSPR, and SAT total score.

Having selected the appropriate covariates, logistic regression was used to estimate the propensity scores because the dichotomous assignment to either the honors or non-honors group served as the outcome variable and the selected covariates were the predictors. ANCOVAs were conducted to estimate the treatment effects in the cohort. For the present study, because (a) the continuous covariates SAT and HSPR and the categorical covariates gender and ethnicity were balanced as a result of the propensity calculation procedure; (b) the nature of the data collection was archival; and (c) the research design was quasi-experimental, the appropriate statistical tests were one-way ANCOVAs (Stangor, 2010) with one between-case factor (i.e., group = honors, non-honors).

Propensity scores for the 2009 cohort were generated for 2,639 cases, with 465 cases excluded because of missing values. For the 2010 cohort, propensity scores were calculated for 2,881 cases, with 573 cases excluded because of missing values. Propensity scores calculated for the 2011 cohort yielded 3,012 cases, with 547 cases excluded because of missing values. Results of the logistic regression analyses indicated that the propensity score four-factor model provided a statistically significant improvement over the constant-only model for each cohort, $\chi^2(9, N = 2639) = 642.19, p < .001$; $\chi^2(4, N = 2881) = 759.58, p < .001$; and $\chi^2(4, N = 3012) = 801.73, p < .001$. The Nagelkerke pseudo-$R^2$ indicated that the model accounted for 48.7%, 46.6%, and 49.1% of the total variance, respectively. These findings suggested that the four-factor model led to a relatively good prediction of participation in the honors college for this sample.

Following the propensity score calculation process, a one-way ANCOVA was used to compare group variance in GPA and magnitude of effect for the treatment group (honors) on first-year cumulative GPA for each cohort.

Participants

Archival data were collected from a large university in the southern United States, for three matriculating cohorts 2009, 2010, and 2011 of full-time, first-time-in-college (FTIC) students. Part-time students and non-FTIC students were excluded because the honor college does not admit students of either status. The office of institutional research provided only data on gender, ethnicity, age, group (honors, non-honors), HSPR, high school quartile, SAT (math, verbal, total), term GPA for each semester enrolled, and cumulative GPA for each spring semester enrolled. High school GPA was not provided. See Table 1 for frequencies and descriptive information.

Table 1. Frequency Table for Demographics.

	Gender		Ethnicity				
Group	Female (n = 4,172)	Male (n = 4,360)	White (n = 2,076)	African American (n = 1,306)	Hispanic (n = 2,380)	Asian (n = 2,460)	Other (n = 380)
Non-honors	43.8	46.3	20.0	14.9	26.3	25.8	3.9
Honors	5.0	4.7	4.3	0.3	1.5	3.0	0.5

Note. Frequencies are presented as percentage of the total sample, N = 8,532. The table includes values only for cases for which propensity scores were calculated.

FTIC students who matriculate into the typical curriculum rather than honors complete 30 college credit courses, including 2 English courses, 2 history courses, 2 political science courses, 2 political science courses, and 2 courses in their major. Unlike the honors courses, typical courses are not capped.

The honors college at the host university offers smaller class sizes (honors courses), accessible faculty/staff, priority registration, reserved facilities and housing, study abroad, and research opportunities. Criteria for admission to the college include: admission to the university, academic record (high school or college), extracurricular activities, test scores, and an essay. The honors college evaluates each applicant on an individual basis, therefore, there is no selection index or cut score on which the selection committee bases its decision to admit. The honors college reports that their average students have graduated in the top 10% of their high school and achieved an SAT score of 1,300 or above.

Honors students complete 36 hr of honors coursework, including core courses, major and minor courses, and electives. Students who successfully complete these courses graduate with the University Honors distinction. Honors students who choose to complete a senior honors thesis graduate with the additional distinction of Honors in the Major. To satisfy honors and core requirements simultaneously, all FTIC honors students complete five additional courses specifically designed for and restricted to honors students.

To control for historical factors such as state financial aid programs, demographic changes, and changes in university admission policies, three separate cohorts were examined. Examining individual cohorts additionally provides a measure of external validity.

Results

There were significant interactions between the covariate (propensity score) and the independent variable group) resulting in a violation of the assumption of homogeneity of regression of slopes, $F(1, 2616) = 28.68, p < .001$; $F(1, 2864) = 15.32, p < .001$; and $F(1, 2994) = 12.95, p < .001$, for 2009, 2010, and 2011, respectively. Appropriate use

of ANCOVA is predicated on the assumption that the covariate is independent of the experimental treatments (Keppel & Wickens, 2004). Some researchers have argued that the use of ANCOVA in quasi-experimental designs remains a controversial issue (Huitema, 2011; Pedhazur, 1997; Tabachnick & Fidell, 2001).

The violation of the assumption of homogeneity of regression means that the intervention (honors) did not affect all participants equally. The effect of the fixed factor (i.e., independent variable [IV] = group) differed depending on the value of the propensity score. Therefore, the results of performing an ANCOVA and reporting adjusted means could have been misleading because they did not convey this important information. When the regression slopes are homogeneous, the adjusted means are appropriate descriptive measures because the treatment effects are constant across all levels of the propensity score (Huitema, 2011). Put simply, this means that the average treatment difference cannot be interpreted as a constant with respect to the propensity score. Therefore, the effect may be greater for larger propensity score values and lesser for smaller values, or the converse. When the difference is small and group sizes are equal, Keppel and Wickens (2004) have argued that this type of heterogeneity is usually not a significant problem and the ANCOVA remains robust. In the present study, the difference is significant and the group sizes are not equal. In this circumstance, as a true treatment–covariate interaction existed, the next step was to determine (a) What is the region of insignificance for group differences when the treatment effect differs depending on the level of the covariate? and (b) For what values of the covariate do the groups differ significantly with regard to the dependent variable? This procedure is referred to as the Johnson–Neyman (J-N) technique (Johnson & Neyman, 1936), in which treatment effects are estimated as a function of the covariate score. Pothoff (1964) and Rogosa (1978) developed further modifications to the J-N procedure.

D'Alonzo (2004), Dorsey and Soeken (1996), and Huitema (2011) advocated the J-N technique in situations such as the present study. The J-N technique may be used if two or more covariates are involved. In the present study, there is only one covariate, but it is derived from four values, two categorical (gender, ethnicity) and two continuous (SAT, HSPR). If the IV consists of three or more groups, Huitema suggested substituting the Bonferroni *F* statistic in place of the conventional *F* statistic. In the present study, the IV consists of only two groups, honors and non-honors, so the Bonferroni correction would not be appropriate. For these reasons, the researcher continued with the analysis using the J-N procedure.

The limits of the region of nonsignificance on *X* are calculated by using

$$X_{L1} = \frac{-B \pm \sqrt{B^2 - AC}}{A}$$

and

$$X_{L2} = \frac{-B \pm \sqrt{B^2 - AC}}{A}$$

Table 2. Distribution of Propensity Scores (Predicted Probabilities).

	2009		2010		2011	
Group	*n*	*M* (*SD*)	*n*	*M* (*SD*)	*n*	*M* (*SD*)
Non-honors	2,412	0.0544 (0.1116)	2,568	0.0713 (0.1319)	2,713	0.0626 (0.1235)
Honors	227	0.4210 (0.2817)	313	0.4149 (0.2669)	299	0.4324 (0.2770)
Total	2,639	0.0860 (0.1696)	2,881	0.1086 (0.1861)	3,012	0.0993 (0.1832)

Note. Propensity scores were calculated using the variables, gender, ethnicity, SAT total score, and HSPR and indicated the predicted probability of participating in the honors program. HSPR = high school percentile ranking.

where X_{L1} and X_{L2} = limits of nonsignificance region,

$$A = \frac{-2F_{(\alpha,1,N-4)}}{n-4}\left(\text{SSres}_1\right)\left(\frac{1}{\sum x_1^2}+\frac{1}{\sum x_2^2}\right)+\left(b_1^{(\text{Group1})}-b_1^{(\text{Group2})}\right)^2,$$

$$B = \frac{2F_{(\alpha,1,N-4)}}{n-4}\left(\text{SSres}_1\right)\left(\frac{1}{\sum x_1^2}+\frac{1}{\sum x_2^2}\right)+\left(b_0^{(\text{Group1})}-b_0^{(\text{Group2})}\right)\left(b_1^{(\text{Group1})}-b_1^{(\text{Group2})}\right),$$

$$C = \frac{-2F_{(\alpha,1,N-4)}}{n-4}\left(\text{SSres}_1\right)\left(\frac{N}{n_1 n_2}+\frac{\bar{X}_1^2}{\sum x_1^2}+\frac{\bar{X}_2^2}{\sum x_2^2}\right)+\left(b_0^{(\text{Group1})}-b_0^{(\text{Group2})}\right)^2,$$

where $F_{(\alpha,\,2,\,N-4)}$ is the critical value of the F statistic for the desired level of α and 1, and N is 4 degrees of freedom (where N is the total number of cases; that is, n_1+n_2); SSres_1 is the individual residual sum of squares; $\bar{X}_0$ and $\bar{X}_1$ are covariate means for Groups 0 and 1, respectively; Σx_0^2 and Σx_1^2 are covariate sums of squares for Groups 0 and 1, respectively; $b_0^{(\text{Group1})}$ is the regression of the intercept for Group 0; $b_0^{(\text{Group2})}$ is the regression of the intercept for Group 1; $b_1^{(\text{Group1})}$ is the regression of the slope for Group 0; and $b_1^{(\text{Group2})}$ is the regression of the slope for Group 1.

To avoid miscalculations by hand, the J-N calculator (Oshima, 2014) was used to compute the regions of nonsignificance for each cohort. Table 2 provides frequency distributions of the propensity scores.

For the 2009 cohort (Table 3), the region of nonsignificance is .2568 and .3558 using an α = .05. Therefore, for students having a propensity score below 0.2568, participation in the honors program is effective, but for individuals who have a propensity score above 0.3558, the intervention appears to have had no effect. For individuals whose propensity scores were between 0.2568 and 0.3558, there is insufficient evidence to conclude that participation in honors program had any effect (Figure 1). For the 2010 cohort (Table 4), the region of nonsignificance is .3716 and .4620 using an α = .05. Therefore, for students having a propensity score below 0.3716, participation in the honors program is effective, but for individuals who have a propensity score above

Table 3. Summary Statistics, Intercepts, Slopes, Critical F values, and Regions of Nonsignificance for 2009 Cohort.

Group	*N*	*M (SD)*	B_0	B_1
Non-honors	2,396	2.58 (0.92)	2.47	1.99
Honors	224	3.15 (0.63)	2.92	0.59

Note. Critical $F(.05, 2, 2616) = 3.00$. Regression lines intersect at $X = 0.3063$. The values for *A, B, C*, and the limits on the region of nonsignificance are $A = 1.96$, $B = -0.60$, $C = 0.18$, $X_{L1} = 0.2568$, and $X_{L2} = 0.3558$.

0.4620, the intervention appears to have had no effect. For individuals whose propensity scores were between 0.3716 and 0.4620, there is insufficient evidence to conclude that participation in honors program had any effect (Figure 2). For the 2011 cohort (Table 5), the region of nonsignificance was .4369 and .5364 using an $\alpha = .05$. Therefore, for students having a propensity score below 0.4369, participation in the honors program is effective, but for individuals who have a propensity score above 0.5364, the intervention appears to have had no effect. For individuals whose propensity scores were between 0.4369 and 0.5364, there is insufficient evidence to conclude that participation in honors program had any effect (Figure 3).

Discussion

As with all quasi-experimental studies and because honors programs vary from institution to institution with regard to structure and student body profile, caution should be employed in generalizing the findings of the present study to other honors programs. Propensity scores indicating students' predicted probability to participate in honors were calculated based on ethnicity, gender, high school percentile ranking, and SAT total score. Propensity scores were then used as the covariate. The results from the present study suggest that the effect of participating in honors on GPA is greater for smaller covariate values and lesser for greater covariate values. Students at the upper end of the propensity score scale demonstrated no effect on the outcome of cumulative GPA as a result of participation in the program.

Although the findings of previous studies indicate that honors programs provide an optimal environment for high-achieving learners (Giazzoni & Hilberg, 2009; Mihelich et al., 2007; Moritz, 2011; Rinn & Cobane, 2009; Shushok, 2002; Sievert, 1972; Spurrier, 2009; Weiner, 2009), the present study suggests that the benefits for participating in the honors program do not accrue for all members of the program as a constant effect. These results provide strong support for encouraging high-achieving students who have a predicted probability of participating in honors yet do not to matriculate into honors, especially those at the lower end of the propensity continuum, to apply for admission to the honors college and to enroll if they are selected. Furthermore, because the results were replicated for each of the three cohorts, internal validity is high.

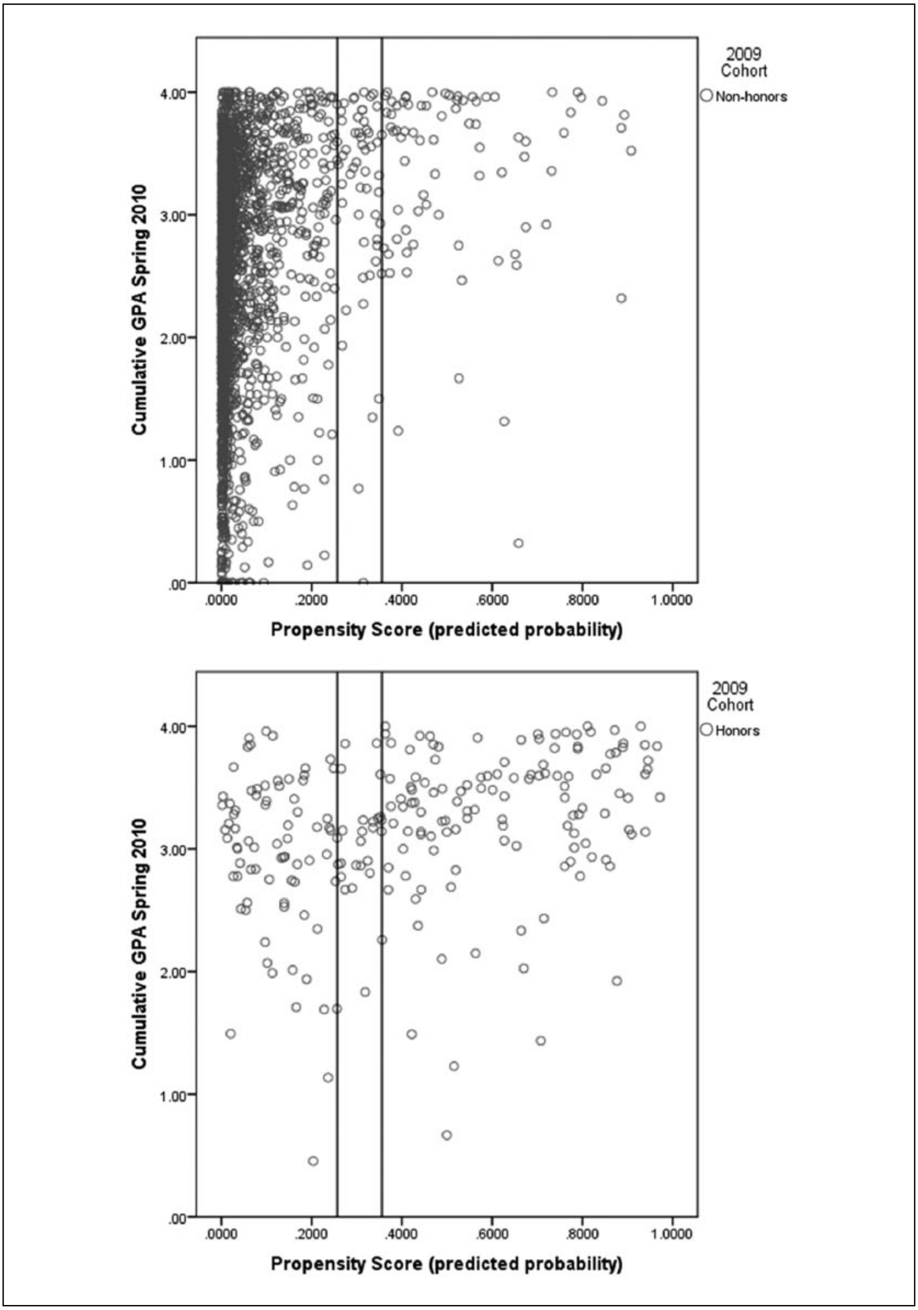

Figure 1. Correlations between cumulative GPA and propensity score for the 2009 cohort.
Note. GPA = grade point average.

Table 4. Summary Statistics, Intercepts, Slopes, Critical F values, and Regions of Nonsignificance for 2010 Cohort.

Group	*N*	*M (SD)*	B_0	B_1
Non-honors	2,558	2.65 (0.93)	2.53	1.165
Honors	310	3.21 (0.64)	2.96	0.62

Note. Critical $F(.05, 2, 2864) = 3.00$. Regression lines intersect at $X = 0.4168$. The values for *A, B, C,* and the limits on the region of nonsignificance are, $A = 1.06$, $B = -0.44$, $C = 0.18$, $X_{L1} = 0.3716$, and $X_{L2} = 0.4620$.

Limitations

The study is limited in several ways, first, by the use of archival data. The sample was not surveyed regarding psychological constructs of motivation, including Grit, academic self-concept, self-efficacy, or achievement goal orientation, all of which have been found to influence academic achievement (Ablard & Lipschultz, 1998; Bandura, 1997; Duckworth, 2006; Elliot, 2005; Moon, 2012; Wouters, Germeijs, Colpin, & Verschueren, 2011). In addition, propensity scores were derived from observable characteristics that included SAT, HSPR, gender, and ethnicity. Characteristics such as socioeconomic status and parents' level of education were not included, both of which may affect academic achievement. Moreover, academic goals and choice of major were not included.

It is altogether appropriate to question the use of GPA as the dependent variable. Honors courses may be subject to grade inflation in addition to being a non-uniform metric of academic performance. Using the covariate adjustment propensity model, the researcher can only adjust for observed variables. For example, high school GPA was not provided to the researcher, therefore, not being an observed variable, it could not be included as a factor in calculating propensity scores. Conversely, including irrelevant characteristics in the propensity model may incur bias.

Although this study was not, by definition, a longitudinal design, it was susceptible to threats of bias and confounders associated with longitudinal study designs, such as meeting the criteria of selective sampling (Baltes, 1968), survival/attrition (Jarvik & Falek, 1963), and selective drop-out/experimental mortality (D. Campbell, Stanley, & Gage, 1963). These threats, however, were minimized in the propensity score method. Despite the advantages of using propensity scores, there remains the issue of fidelity to treatment. Both the honors and the non-honors groups were enrolled in a variety of courses and exposed to a variety of educational tasks and learning strategies. Furthermore, the professional experience and level of expertise possessed by the faculty who taught these students varied widely. Other confounders such as the lack of uniformity in the measure of GPA and the possibility of grade inflation by honors instructors have not been included as factors. Lack of control over these aspects threatens the validity of the findings.

Future research examining the effect of participating in a post-secondary honors program would benefit by including several variables, such as measures of Grit, academic self-concept, self-efficacy, achievement goal orientation, parental involvement in previous education, socioeconomic status, and parents' level education. What particular interaction of factors within the honors context that influences this outcome was beyond the scope of this study, but would be worthy of future exploration. Although

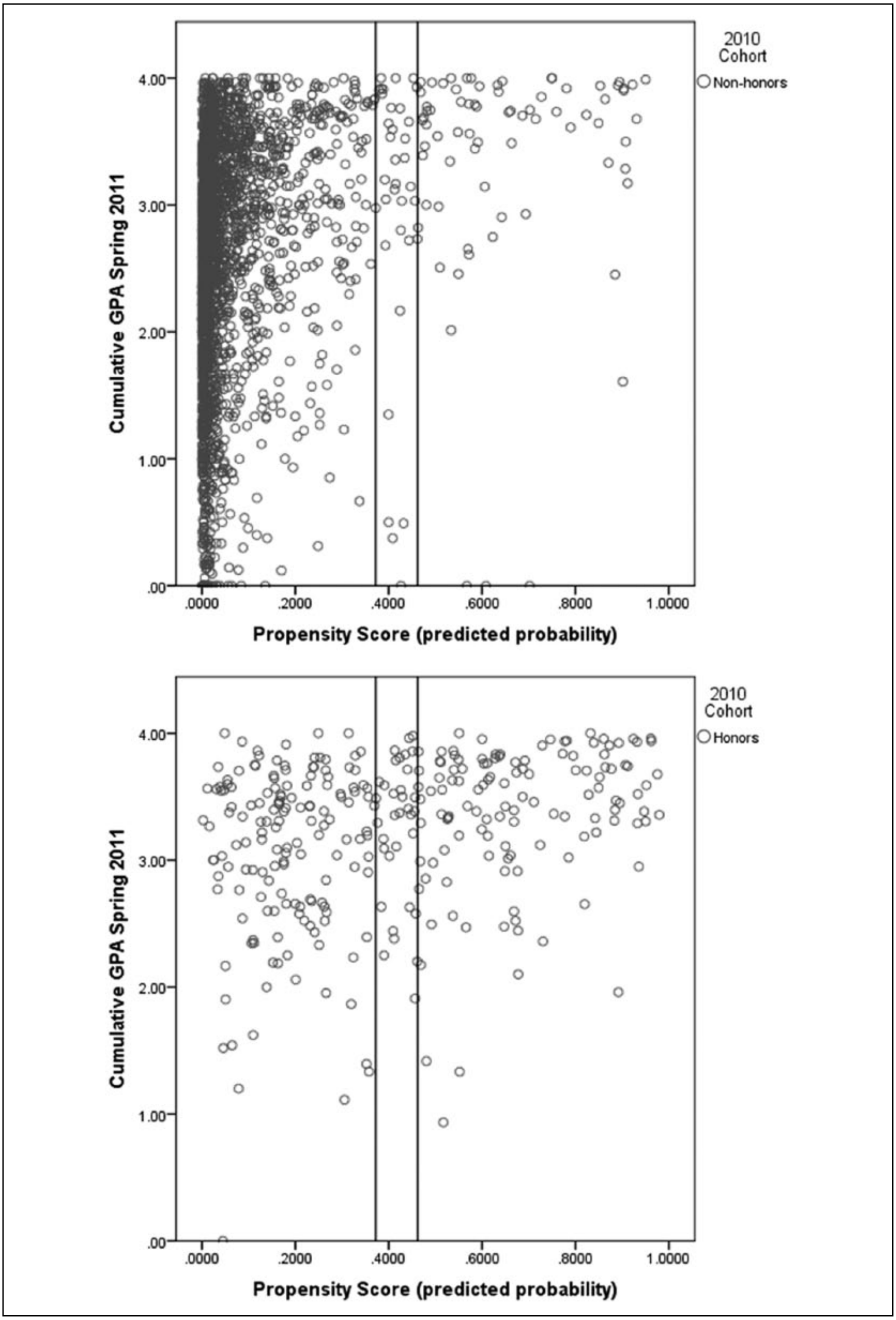

Figure 2. Correlations between cumulative GPA and propensity score for the 2010 cohort.
Note. GPA = grade point average.

Table 5. Summary Statistics, Intercepts, Slopes, Critical F values, and Regions of Nonsignificance for 2011 Cohort.

Group	*N*	*M (SD)*	B_0	B_1
Non-honors	2,700	2.73 (0.90)	2.64	1.35
Honors	298	3.27 (0.58)	3.03	0.55

Note. Critical $F(.05, 2, 2994) = 3.00$. Regression lines intersect at $X = 0.4867$. The values for *A, B, C*, and the limits on the region of nonsignificance are, $A = 0.64$, $B = -0.31$, $C = 0.15$, $X_{L1} = 0.4369$, and $X_{L2} = 0.5364$.

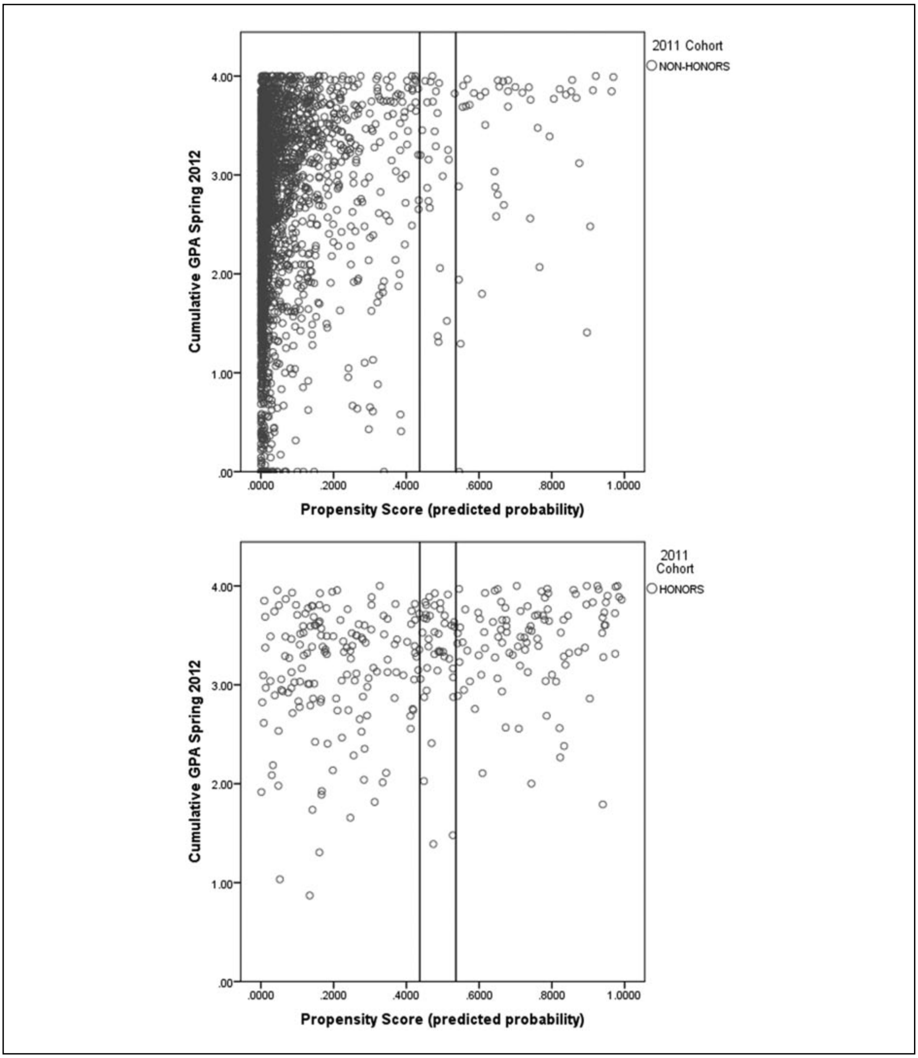

Figure 3. Correlations between cumulative GPA and propensity score for the 2011 Cohort.
Note. GPA = grade point average.

true experimental design, which incorporates random selection and assignment of participants, is unlikely in a field study, it may be possible to identify an institution where all honors students and non-honors students undertake a similar curriculum for better control of confounders related to varied curricula and instruction.

Given the number of high-achieving students who attend large public universities, it would be important to explore the effect of honors and non-honors participation on course success, course completion, retention, and time to graduation. In addition, future research could examine the interaction between gender, ethnicity, and group (honors and non-honors) on academic achievement, among other academic outcomes. Given the differences between the two groups on the characteristics of gender and ethnicity, it would be beneficial to understand the relationship between gender, ethnicity, and their influence on decisions to participate and thrive in a post-secondary honors program.

In addition to gender and ethnicity, programmatic characteristics such as scholarships, criteria to remain in honors, living-learning communities, peer effects, teacher experience, teachers' perceptions and expectations of honors students, class-size, and designated honors facilities should be taken into consideration as variables that could potentially influence the effect size for participation in honors. The prospective ramifications of such research could be instrumental in the student success of high-achieving students who are at risk of stopping out or dropping out altogether. Universities are being held more accountable of creating an environment and providing resources not only to recruit the best and the brightest but also to optimize their likelihood of graduating. Because students are high-achieving does not mean they are necessarily equipped to handle successfully the transition from high school to college.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Ablard, K., & Lipschultz, R. (1998). Self-regulated learning in high-achieving students: Relations to advanced reasoning, achievement goals, and gender. *Journal of Educational Psychology*, *90*, 94-101. doi:10.1037/0022-0663.90.1.94

Achterberg, C. (2005). What is an honors student? *Journal of the National Collegiate Honors Council*, *6*, 75-83.

Astin, A. (1978). *Four critical years*. San Francisco, CA: Jossey-Bass.

Astin, A. (1993). *What matters in college?* San Francisco, CA: Jossey-Bass.

Austin, P. (2011). An introduction to propensity score methods for reducing effects of confounding in observational studies. *Multivariate Behavioral Research*, *46*, 300-424.

Bacon, D., & Bean, B. (2006). GPA in research studies: An invaluable but neglected opportunity. *Journal of Marketing Education*, *28*, 35-42. doi:10.1177/0273475305284638

Baltes, P. (1968). Longitudinal and cross-sectional sequences in the study of age and generation effects. *Human Development*, *11*, 145-171.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.
Bloom, B. (1976). *Human characteristics and school learning*. New York, NY: McGraw-Hill.
Blythe, H. (2004). Ethics on an honors college campus: An analysis of attitudes and behaviors of honors versus non-honors students. *Journal of the National Collegiate Honors Council*, *5*, 25-35.
Borde, S. (1998). Predictors of student academic performance in the introductory marketing course. *Journal of Education for Business*, *73*, 302.
Burke, J., & Modarresi, S. (2001). Performance funding programs: Assessing their stability. *Research in Higher Education*, *42*, 51-70.
Butler, J., Pryor, B., & Marti, S. (2004). Communication apprehension and honors students. *North American Journal of Psychology*, *6*, 293-296.
Campbell, D., Stanley, J., & Gage, N. (1963). *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton Mifflin.
Campbell, K. (2005). Allocation of resources: Should honors programs take priority? *Journal of the National Collegiate Honors Council*, *6*, 95-103.
Castro-Johnson, M., & Wang, A. (2003). Emotional intelligence and academic performance of college honors and non-honors freshmen. *Journal of the National Collegiate Honors Council*, *4*, 105-114.
Chapman, K., Davis, R., Toy, D., & Wright, L. (2004). Academic integrity in the business school environment: I'll get by with a little help from my friends. *Journal of Marketing Education*, *26*, 236-249.
Chase, C., Hemmeter, J., & Griffin, N. (1971). *Grade point averages: Honors vs. non-honors courses* (Indiana Studies in Prediction, No. 15). Bloomington: Bureau of Educational Studies and Testing, Indiana University.
Cosgrove, J. (2004). The impact of honors programs on undergraduate academic performance, retention, and graduation. *Journal of the National Collegiate Honors Council*, *5*, 45-53.
D'Alonzo, K. (2004). The Johnson-Neyman procedure as an alternative to ANCOVA. *West Journal of Nursing Research*, *26*, 804-812.
Dorsey, S., & Soeken, K. (1996). Use of the Johnson-Neyman procedure as an alternative to analysis of covariance. *Nursing Research*, *45*, 363-366.
Duckworth, A. (2006). Intelligence is not enough: Non-IQ predictors of achievement. *Dissertation Abstracts International*, *67*, 1741.
Elliot, A. (2005). A conceptual history of the achievement goal construct. In A. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 52-72). New York, NY: Guilford.
Giazzoni, M., & Hilberg, N. (2009). Dealing with subjective and objective issues in honors education. *Journal of the National Collegiate Honors Council*, *10*, 57-59.
Hartleroad, G. E. (2005). Comparison of the academic achievement of first-year female honors program and non-honors program engineering students. *Journal of the National Collegiate Honors Council*, *6*, 109-120.
Hébert, T., & McBee, M. (2007). The impact of an undergraduate honors program on gifted university students. *Gifted Child Quarterly*, *51*, 136-151.
Hite, R., McIntyre, F., & Lynch, D. (2001). The impact of student characteristics on cooperative testing in the marketing classroom. *Marketing Education Review*, *11*, 27-34.
Hoxby, C., & Avery, C. (2012). *The missing "one-offs": The hidden supply of high-achieving, low income students* (Working Paper No. 18586). Washington, DC: National Bureau of Economic Research.

Huitema, B. (2011). *The analysis of covariance and alternatives: Statistical methods for experiments, quasi-experiments, and single-case studies*. Hoboken, NJ: Wiley.

Jaramillo, F., & Spector, P. (2004). The effect of action orientation on the academic performance of undergraduate marketing majors. *Journal of Marketing Education*, *26*, 250-260. doi:10.1177/0273475304268780.

Jarvik, L., & Falek, A. (1963). Intellectual stability and survival in the aged. *Journal of Gerontology*, *18*, 173-176.

Johnson, P., & Neyman, J. (1936). Tests of certain linear hypotheses and their applications to some educational problems. *Statistical Research Memoirs*, *1*, 57-93.

Karakaya, F., Ainscough, T., & Chopoorian, J. (2001). The effects of class size and learning style on student performance in a multimedia-based marketing course. *Journal of Marketing Education*, *23*, 84.

Keller, R., & Lacy, M. (2013). Propensity score analysis of an honors program's contribution to students' retention and graduation outcomes. *Journal of the National Collegiate Honors Council*, *13*, 73-84.

Keppel, G., & Wickens, D. (2004). *Design and analysis: A researcher's handbook*. Upper Saddle River, NJ: Pearson.

McDermott, D. (1989). College honors programs foster elitism and threaten morale. *Chronicle of Higher Education*, *35*, B2.

Mihelich, J., Storrs, D., & Pellett, P. (2007). Transformational experience through liberation pedagogy: A critical look at honors education. *Journal of the National Collegiate Honors Council*, *8*, 97-115.

Moon, J. (2012). *Honors and high-ability students: Factors that predict academic efficacy, critical thinking skills, and academic goals*. Available from Graduate Theses and Dissertations. (Paper 12412)

Moritz, B. (2011). Can the elitism of honors help students at non-elite schools? *Journal of the National Collegiate Honors Council*, *12*, 65-67.

Morris, M. (2008). Examining teachers' beliefs and instructional practices in secondary foreign language classes on honors and non-honors tracks: The role of teaching for communication. *Teachers College Record*, *110*, 1065-1087.

Ogilvie, K., & Reza, E. (2009). Business student performance in traditional vs. honors course settings. *Business Education Innovation Journal*, *1*, 31-37.

Oshima, T. (2014). *QJN.xlsx*. Faculty website. Department of Educational Policy Studies, College of Education, Georgia State University. Retrieved from http://coeweb.gsu.edu/coshima/stat3.htm

Pedhazur, E. (1997). *Multiple regression in behavioral research: Explanation and prediction* (2nd ed.). New York, NY: Hold, Rinehart and Winston.

Pehlke, J. (2003). The myth of an honors education. *Journal of the National Collegiate Honors Council*, *4*, 27-33.

Pothoff, R. (1964). On the Johnson-Neyman technique and some extensions thereof. *Psychometrika*, *29*, 241-256.

Rice, K., Leever, B., Christopher, J., & Porter, J. (2006). Perfectionism, stress, and social (dis)connection: A short-term study of hopelessness, depression, and academic adjustment among honors students. *Journal of Counseling Psychology*, *53*, 524-534.

Rinn, A. (2005). Trends among honors college students: An analysis by year in school. *Journal of Secondary Gifted Education*, *16*, 157-167.

Rinn, A. (2007). Effects of programmatic selectivity on the academic achievement, academic self-concepts, and aspirations of gifted college students. *Gifted Child Quarterly*, *51*, 232-245.

Rinn, A., & Cobane, C. (2009). Elitism misunderstood: In defense of equal opportunity. *Journal of the National Collegiate Honors Council*, *10*, 53-56.

Rogosa, D. (1978). *Some results for the Johnson-Neyman procedure. Dissertation Abstracts International*, *38*, 5366.

Roth, P., & Clarke, R. (1998). Meta-analyzing the relationship between grades and salary. *Journal of Vocational Behavior*, *53*, 386-400.

Seifert, T., Pascarella, E., Colangelo, N., & Assouline, S. (2007). The effects of honors program participation on experiences of good practices and learning outcomes. *Journal of College Student Development*, *48*, 57-74.

Selingo, J. (2002). Mission creep? *Chronicle of Higher Education*, *48*, A19-A21.

Shadish, W., Clark, M., & Steiner, P. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *Journal of the American Statistical Association*, *103*, 1334-1344.

Shushok, F. (2002). *Educating the best and the brightest: Collegiate honors programs and the intellectual, social and psychological development of students* (Doctoral dissertation). University of Maryland, College Park, MD.

Shushok, F. (2006). Student outcomes and honors programs: A longitudinal study of 172 honors students, 2000-2004. *Journal of the National Collegiate Honors Council*, *7*, 85-96.

Siegle, D., Rubenstein, L., Pollard, E., & Romey, E. (2010). Exploring the relationship of college freshmen honors students' effort and ability attribution, interest, and implicit theory of intelligence with perceived ability. *Gifted Child Quarterly*, *54*, 92-101.

Sievert, W. (1972). College honors programs, accused of elitism, facing crises. *Chronicle of Higher Education*, *7*, 1-5.

Slavin, C., Coladarci, T., & Pratt, P. (2008). Is student participation in an honors program related to retention and graduation rates? *Journal of the National Collegiate Honors Council*, *66*, 59-69.

Spurrier, R. (2009). To the charge of "honors is elitist," on advice of counsel we plead "guilty as charged." *Journal of the National Collegiate Honors Council*, *10*, 49-51.

Stangor, C. (2010). *Research methods for the behavioral sciences*. Belmont, CA: Cengage Learning.

Tabachnick, B., & Fidell, L. (2001). *Using multivariate statistics*. Boston, MA: Allyn & Bacon.

Tam, M., & Sukhatme, U. (2003). The importance of high school quality in university admissions decisions. *College & University*, *78*, 3-8.

Tam, M., & Sukhatme, U. (2004). How to make better college admissions decisions: Considering high school quality and other factors. *The Journal of College Admission*, *183*, 12-16.

Teske, B., & Etheridge, B. (2010). Information and communication technology literacy among first-year honors and non-honors students: An assessment. *Journal of the National Collegiate Honors Council*, *11*, 83-109.

Thoemmes, F., & Kim, E. (2011). A systematic review of propensity score methods in the social sciences. *Multivariate Behavioral Research*, *46*, 90-118.

Weiner, N. (2009). Honors is elitist, and what's wrong with that? *Journal of the National Collegiate Honors Council*, *10*, 19-24.

What Works Clearinghouse. (2011). *Procedures and Standards Handbook* (Version 2.1). Washington, DC: U. S. Department of Education.

Wouters, S., Germeijs, V., Colpin, H., & Verschueren, K. (2011). Academic self-concept in high school: Predictors and effects on adjustment in higher education. *Scandinavian Journal of Psychology*, *52*, 586-594.

Wurst, C., Smarkola, C., & Gaffney, M. (2008). Ubiquitous laptop usage in higher education: Effects on student achievement, student satisfaction, and constructivist measures in honors and traditional classrooms. *Computers & Education*, *51*, 1766-1783.